

An optimal bound on the quantiles of a certain kind of distributions

Iosif Pinelis*

*Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu*

Abstract: An optimal bound on the quantiles of a certain kind of distributions is given. Such a bound is used in applications to Berry–Esseen-type bounds for nonlinear statistics.

AMS 2010 subject classifications: Primary 60E15; secondary 62E17.

Keywords and phrases: quantiles, distributions, optimal bounds, moments, Berry–Esseen bounds, probability inequalities, nonlinear statistics.

Let μ be any probability measure μ on $(0, \infty)$. For any real p , let

$$\mu_p := \int_{(0, \infty)} x^{p-2} \mu(dx). \quad (1)$$

Consider the function $L: (0, \infty) \rightarrow \mathbb{R}$ defined by the formula

$$L(d) := L_\mu(d) := \int_{(0, \infty)} (1 \wedge \frac{d}{x}) \mu(dx).$$

Clearly, L is continuous and nondecreasing, with $L(0) = 0$ and $L(\infty-) = 1$.

Take now an arbitrary $c \in (0, 1)$. Then the equation

$$L(d) = c \quad (2)$$

has a root $d \in (0, \infty)$. Moreover, this root is unique. Indeed, if $L(d) = c$ for some $d \in (0, \infty)$, then $\int_{(d, \infty)} \frac{x-d}{x} \mu(dx) = L(\infty-) - L(d) = 1 - c > 0$ and hence $\int_{(d, \infty)} \mu(dx) > 0$ and the right derivative of the function of L at the point d is $\int_{(d, \infty)} \frac{1}{x} \mu(dx) > 0$. So, the definition

$$\delta := \delta_\mu := L^{-1}(c) = L_\mu^{-1}(c) \quad (3)$$

is proper. Note that δ may be viewed as the c -quantile of the distribution function L of a probability distribution on $(0, \infty)$.

Theorem 1. *If $\mu_3 < \infty$, then*

$$\delta \leq \delta_* := \begin{cases} c\mu_3 & \text{if } 0 < c \leq \frac{1}{2}, \\ \frac{\mu_3 - (2c-1)^2/\mu_1}{4(1-c)} & \text{if } \frac{1}{2} \leq c < 1. \end{cases} \quad (4)$$

*Supported in part by NSA grant H98230-12-1-0237

Note that the two expressions for δ_* in (4) in the case $c = \frac{1}{2}$ have the same value, $\mu_3/2$.

Proof of Theorem 1. Consider first the case $\frac{1}{2} < c < 1$. Take any real numbers d and u such that $0 < u < d$ and introduce also the functions f and g on $(0, \infty)$ defined by the formulas

$$f(x) := (1 \wedge \frac{d}{x}) - c \quad \text{and} \quad g(x) := b_0 - b_3x - \frac{b_1}{x}$$

for $x > 0$, where

$$b_0 := \frac{2(1-c)d + (2c-1)u}{2(d-u)}, \quad b_3 := \frac{1}{4(d-u)}, \quad b_1 := \frac{u^2}{4(d-u)};$$

note that $b_1 > 0$, so that g is strictly concave on $(0, \infty)$. Let also

$$v := 2d - u. \tag{5}$$

Then $v \in (d, \infty)$, and one can check that $(f - g)(w) = (f - g)'(w) = 0$ for $w \in \{u, v\}$. Since the function $f - g$ is strictly convex on $(0, d]$ and on $[d, \infty)$, it follows that $f > g$ on $(0, \infty) \setminus \{u, v\}$ and $f = g$ on $\{u, v\}$. So,

$$\begin{aligned} L(d) - c &= \int_{(0, \infty)} f \, d\mu \\ &\geq \int_{(0, \infty)} g \, d\mu = b_0 - b_3\mu_3 - b_1\mu_1 = \frac{1-c}{d-u} (d - d_*(u)) = 0 \end{aligned} \tag{6}$$

if $d = d_*(u)$, where

$$d_*(u) := \frac{\mu_3 - 2u(2c-1) + \mu_1 u^2}{4(1-c)}. \tag{7}$$

Next,

$$d_*(u) \geq d_*(u_*) = \delta_*,$$

where

$$u_* := \frac{2c-1}{\mu_1}. \tag{8}$$

Obviously, $u_* > 0$. Also, μ_p (defined by (1)) is log-convex in $p > 0$ and hence

$$\mu_3\mu_1 \geq \mu_2^2 = 1. \tag{9}$$

So, $\delta_* - u_* = \frac{4(1-c)^2 + \mu_1\mu_3 - 1}{4(1-c)\mu_1} > 0$, and hence $0 < u_* < \delta_*$. Thus, $L(\delta_*) = L(d_*(u_*)) \geq c$, and the inequality $\delta \leq \delta_*$ in (4) in the case $\frac{1}{2} < c < 1$ follows by the monotonicity of the function L .

The case $0 < c \leq \frac{1}{2}$ is similar and even simpler. Take here $d = \delta_* = c\mu_3$ and $g(x) := c - cx/\mu_3$ for $x > 0$. Then $(f - g)(0+) = 1 - 2c \geq 0$ and $(f - g)(\mu_3) = (f - g)'(\mu_3) = 0$. Since the function $f - g$ is strictly convex on $[d, \infty)$ and affine on $(0, d]$, it follows that $f > g$ on $(0, \infty) \setminus \{\mu_3\}$ and $f = g$ on $\{\mu_3\}$. So,

$$L(c\mu_3) - c = \int_{(0, \infty)} f \, d\mu \geq \int_{(0, \infty)} g \, d\mu = 0, \tag{10}$$

and the inequality $\delta \leq \delta_*$ in (4) in the case $0 < c \leq \frac{1}{2}$ follows by the monotonicity of L . \square

Remark 2. It is clear from the above proof of Theorem 1 that the inequality $\delta \leq \delta_*$ in (4) is strict unless the support of the measure μ consists of one point (in the case $0 < c \leq \frac{1}{2}$) or of two points (in the case $\frac{1}{2} < c < 1$).

Moreover, the upper bound δ_* on δ is the best possible one in terms of c , μ_3 , and μ_1 in the following sense:

Proposition 3.

- (I) For any $c \in (0, \frac{1}{2}]$ and any positive real number μ_{3*} , there exists a probability measure μ on $(0, \infty)$ such that $\delta = \delta_*$ and (1) holds for $p = 3$ with μ_{3*} in place of μ_3 .
- (II) For any $c \in (\frac{1}{2}, 1)$ and any positive real numbers μ_{3*} and μ_{1*} such that $\mu_{3*}\mu_{1*} \geq 1$ (cf. (9)), there exists a probability measure μ on $(0, \infty)$ such that $\delta = \delta_*$ and (1) holds for $p \in \{1, 3\}$ with μ_{3*} and μ_{1*} in place of μ_3 and μ_1 .

Proof of Proposition 3. Let us consider first the more complicated case (II).

(II). Take indeed any $c \in (\frac{1}{2}, 1)$ and any positive real numbers μ_{3*} and μ_{1*} such that $\mu_{3*}\mu_{1*} \geq 1$. Let δ_* and u_* be defined as in (4) and (8), respectively, but with μ_{3*} and μ_{1*} in place of μ_3 and μ_1 . As shown in the proof of Theorem 1, $0 < u_* < \delta_*$. Now, in accordance with (5) and (7), introduce $v_* := 2\delta_* - u_*$; it follows that $v_* > \delta_*$. Let μ be the probability measure with masses $1 - \pi$ and π at the points u_* and v_* , respectively, where $\pi := \frac{2(1-c)(\mu_{3*}\mu_{1*} - (2c-1))}{4(1-c)^2 + \mu_{3*}\mu_{1*} - 1}$; note that such a measure μ exists, since $0 < \pi \leq 1$. Moreover, one can check that then (1) holds for $p \in \{1, 3\}$ with μ_{3*} and μ_{1*} in place of μ_3 and μ_1 . Recalling now that $f = g$ on the set $\{u, v\}$ and using (6) with $\delta_* = d_*(u_*)$, u_* , μ_{3*} , and μ_{1*} in place of d , u , μ_3 , and μ_1 , one concludes that $\delta_* = \frac{1}{4(1-c)}(\mu_{3*} - \frac{(2c-1)^2}{\mu_{1*}})$ is indeed a positive real root d of the equation (2). Finally, the uniqueness of such a root was established in the paragraph containing (3).

(I). The case $c \in (0, \frac{1}{2}]$ is similar but simpler. Here we let μ be the Dirac probability measure with mass 1 at the point μ_3 . Then the inequality in (10) turns into the equality. \square

Take now any natural n and let ξ_1, \dots, ξ_n be any random variables such that $\mathbb{E} \xi_1^2 + \dots + \mathbb{E} \xi_n^2 = 1$. Let then μ_ξ be the probability measure on $(0, \infty)$ defined by the condition $\int_{(0, \infty)} h d\mu = \sum_{i=1}^n \mathbb{E} h(|\xi_i|) \xi_i^2$ for all (say) nonnegative Borel functions h on $[0, \infty)$.

For $\mu = \mu_\xi$ and the “median” value $c = \frac{1}{2}$, the upper bound $\delta_* = \mu_3/2$ on δ follows immediately from the inequality due to Chen and Shao [1, Remark 2.1], who showed that

$$\delta \leq \left(\frac{2(p-2)^{p-2}}{(p-1)^{p-1}} \mu_p \right)^{\frac{1}{p-2}} \quad (11)$$

for $p > 2$. On the other hand, the bound δ_* in (4) is more general than the one in (11) in the sense that c in (4) is allowed to take any value in the interval $(0, 1)$; this flexibility allows one to improve the corresponding results in [2].

References

- [1] L. H. Y. Chen and Q.-M. Shao. Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli*, 13(2):581–599, 2007.
- [2] I. Pinelis and R. Molzon. Berry-Esséen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s (preprint), arXiv:0906.0177v1 [math.ST].